DOCUMENT RESUME

ED 267 247 CE 043 807

AUTHOR Sewall, Timothy J.
TITLE The Measurement of Learning Style: A Critique of Four Assessment Tools.
INSTITUTION Wisconsin Univ., Green Bay. Assessment Center.
PUB DATE Feb 86
NOTE 70p.
PUB TYPE Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS Adult Education; *Adult Learning; *Cognitive Style; Personality Measures; Postsecondary Education; *Response Style (Tests); Student Evaluation; *Test Format; Test Reliability; *Test Selection; *Test Use; Test Validity
IDENTIFIERS Canfield Learning Styles Inventory; Gregorc Style Delineator; Kolb Learning Style Inventory; Myers Briggs Type Indicator

ABSTRACT

This paper addresses the issue of whether four of the learning styles instruments currently available are of sufficient psychometric quality to warrant their continued use either for research or educational purposes. Four instruments, which purport to measure learning styles, were selected for review. Criteria for selection were based in part on the frequency of references in the professional literature and discussions with several adult educators who have had considerable experience with the assessment of learning styles. The four instruments chosen were the Myers-Briggs Type Indicator, the Kolb Learning Style Inventory, Canfield's Learning Styles Inventory, and Gregorc's Style Delineator. The paper is divided into six sections. Following the first section, which introduces the study, each of the next four sections consists of a detailed critique of the four learning styles instruments selected for review. Each critique follows the same format and includes a description of (1) the practical features of the test, administration, scoring, and other considerations; (2) characteristics of the manual, including how information is reported and what test interpretation information is provided; (3) characteristics of the test including norms, reliability, validity, and its overall quality; and (4) a summary statement that focuses on the reviewer's personal decision regarding the use of the test. The paper closes with a listing of research questions that need to be explored and some suggestions for improving the measurement of learning styles. (KC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED267247

THE MEASUREMENT OF LEARNING STYLE: A CRITIQUE OF FOUR ASSESSMENT TOOLS

Timothy J. Sewall

Wisconsin Assessment Center
University of Wisconsin-Green Bay
Green Bay, Wisconsin 54301-7001

February 1986

CE043807

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.

Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

T J Sewall

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TABLE OF CONTENTS

INTRODUCTION 1

Definitions of Learning Style 2

Measurement of Learning Styles 3

Purpose of this Paper 4

Instruments Reviewed 5

THE MYERS-BRIGGS TYPE INDICATOR 8

Practical Features of the Test
Administration 8
Scoring 8
Other Considerations 9

Characteristics of the Manual
Reporting of Information 9
Test Intepretation. 10
Technical Information Provided 10

Characteristics of the Test
Norms 11
Reliability 12
Validity 14
Overall Quality of the MBTI 16

Personal Decision Regarding Use of the MBTI 17

KOLB LEARNING STYLE INVENTORY 19

Practical Features of the Test
Administration 19
Scoring 19
Other Considerations 20

Characteristics of the Manual
Reporting of Information 21
Test Intepretation. 21
Technical Information Provided 22

Characteristics of the Test
Norms 23
Reliability 25
Validity 28
Overall Quality of the LSI 34

Personal Decision Regarding Use of the LSI 35

CANFIELD LEARNING STYLES INVENTORY 37

Practical Features of the Test
Administration 37
Scoring 38
Other Considerations 38

Characteristics of the Manual
Reporting of Information 39
Test Intepretation. 39

Characteristics of the Test
Norms 40
Reliability 41
Validity 42
Overall Quality of the LSI 43

Personal Decision Regarding Use of the LSI 44

GREGORC STYLE DELINEATOR 45

Practical Features of the Test
Administration 45
Scoring 45
Other Considerations 46

Characteristics of the Manual
Reporting of Information 46
Test Intepretation. 47

Characteristics of the Test
Norms 48
Reliability 49
Validity 51
Overall Quality of the Delineator 52

Personal Decision Regarding Use of the Delineator 53

SUMMARY, CONCLUSIONS AND SUGGESTIONS FOR FURTHER RESEARCH 54

Summary 54
Content, Format and Scoring 54
Reliability 55
Validity 56

Conclusions 56

Suggestions for Further Research 59

INTRODUCTION

For some time, educators have recognized that learners have different ways of collecting and organizing information into useful knowledge. Correspondingly, not everyone can benefit to the same extent from the same method of instruction. One way educators have found to address the problem of multimethod learning has been to "individualize instruction" a buzzword for educators at all academic levels, who are concerned with tailoring instructional approaches to the needs, interests and skill levels of the learner. Recently, educators, looking for a "scientific" way to determine how learners learn best, have turned to learning style theory to provide a better match between how a person best gains knowledge and the methods used to impart that knowledge.

In 1979, the National Association of Secondary School Principal's director of research, Jim Keefe, wrote:

> "Learning style diagnosis opens the door to placing individualized instruction on a more rational basis. It gives the most powerful leverage yet available to educators to analyze, motivate, and assist students in school. As such, it is the foundation of a truly modern approach to education." (1979, p. 132)

In the last decade, a number of people have developed applied models that use the concept of learning styles. And there have been numerous scales and instruments designed to measure individual differences in learning style (e.g. Canfield and Lafferty, 1974; Gregorc, 1984; Kolb, 1976). In addition, some educators who have made curricular adjustments have reported success with learning style based instruction (Dunn, 1981; Jenkins, 1982; Pizzo, 1982).

While there has been a great deal of interest in the learning style

concept, the measurement of learning styles and the educational application of learning style information is a relatively recent educational phenomenon. In fact, the educational application of this construct is so new, there is still a lack of concensus regarding some basic issues pertaining to learning style. For example, how can learning style best be defined? What is the most appropriate way to measure learning styles? What are the basic components of learning style?

Definition of Learning Style

In general terms, learning style refers to an individual's unique way of interacting with the environment. It is a hypothetical construct that is intended to help explain the learning process. Keefe (1979) suggests that, "Learning styles are characteristic cognitive, affective, and physiological behaviors that serve as relatively stable indicators of how learners perceive, interact with, and respond to the learning environment" (p. 4). In addition, most researchers and educators treat the term "learning style" as a generic term to include the concepts of cognitive style and student response style. Claxton and Ralston (1978) use the term learning style to refer to a "student's consistent way of responding to and using stimuli in the context of learning" (p. 7). In their review of the ERIC literature, research on learning styles was divided into three sections, cognitive style, student response styles and integrated models of learning styles.

Smith (1982) contends that learning style has three major components: the individualized cognitive, affective, and environmental factors. Cognitive factors include field-independence versus field-dependence, a concept formulated by Herman Witkin and his associates (Witkin & Goodenough, 1981); conceptualizing and categorizing which is based on the work of Kagan and Kogan (1970), Kolb and Fry (1975) and others (e.g. Messick, 1984); reflectivity

versus impulsivity as measured by the Matching Familiar Figures Test (O'Donnell, Paulson and McGann, 1978) and an individual's relative reliance on the respective senses for experiencing and organizing information.

Affective considerations include the amount of structure and authority the learner prefers, expectations and motivation, and the degree of interest in the subject matter to be learned. Finally, environmental factors can range from very specific things such as preferred room temperature to the amount of emotional support learners need in the immediate learning environment.

But perhaps one of the most descriptive statements of learning style can be found in Smith's (1982) Learning to Learn, when he asks:

> "What do we mean by stlye? It has long been apparent to teachers, educators, and observers that people differ in how they go about certain activities associated with learning. They differ as to how they approach problem solving. They differ as to how they go about "information processing", or putting information through their minds. Some people like to "get the big picture" of a subject first and then build to a full understanding of that picture by details and examples. Other people like to begin with examples and details and work through to some kind of meaningful construct or way of looking at an area of knowledge out of these details. Some like theory before going into practice. Others don't." (p. 23.)

Measurement of Learning Styles

An examination of the recent research literature pertaining to educational applications of learning style concepts suggests that educators have made a concerted effort to bridge the gap between theory and practice. To a large extent, they have based their investigations on the work of Herman Witkin and others who have done considerable research on cognitive style. For example,

there have been hundreds of articles, book chapters, etc., based on Witkin's field-dependent/field-independent construct. While much of this research has incorporated experimental designs, few experiments related directly to educational issues.

It is interesting to note that most of the learning style literature is based on the results of the cognitive style research. Furthermore, many researchers make an a priori assumption that learning style is measurable (e.g. Cross, 1976; Keefe, 1979) and that the instruments used do provide a valid measure of the learning style construct. Educators and researchers have used teaching style to address a wide range of educational issues such as matching and changing styles and modifying instructional and counseling approaches. In most instances, however, there has been little attention directed toward the questions of how reliabile and valid the instruments are.

Purpose of this Paper

This paper addresses the issue of whether four of the learning styles instruments currently available are of sufficient psychometric quality to warrant their continued use either for research or educational purposes. To what extent do the tests measure what they are intended to measure? Are the results consistent across time? How are the scores derived? Does the standardization sample adequately represent adult student populations? Is sufficient information provided by the publisher to judge the quality of the instrument?

Four instruments, which purport to measure learning styles, were selected for review. The criteria for selection was somewhat arbitrary but was based in part on the frequency of references in the professional literature and discussions with several adult educators who have had considerable experience with the assessment of learning styles.

Instruments

The four instruments chosen were the Myers-Briggs Type Indicator (Myers, 1962), the Kolb Learning Style Inventory (Kolb, 1976), Canfield's Learning Styles Inventory (Canfield, 1980) and Gregorc's Type Indicator (Gregorc, 1984). While all the instruments are self-administered, paper-and-pencil tests, each approaches the measurement of learning style from a slightly different perspective and theoretical base. Figure I adapted from Dunn DeBello, Brennan and Murrain (1981) provides a brief description of the theoretical basis and the major applications of the four instruments. The chart is intended to serve as a reference for individuals interested in a quick overview of these instruments.

The remaining portion of this paper is divided into five sections. Each of the next four sections consists of a detailed critique of the learning styles instruments selected for review. Each critique follows the same format and includes a description of (1) the practical features of the test, administration, scoring and other considerations; (2) characteristics of the manual, including how information is reported and what test interpretation information is provided; (3) characteristics of the test including norms, reliability, validity and its overall quality, and (4) a summary statement which focuses on this reviewer's personal decision regarding the use of the test.

The paper closes with a listing of research questions which need to be explored and some suggestions for improving the measurement of learning styles.

Figure 1. A Comparison of Learning Styles Instruments[1]

The Myers Briggs Type Indicator

Definition of Learning Style: Learners are orderly and consistent in the way that they use perception and judgement. Perception includes the processes of becoming aware of things, people or ideas. Judgement includes the processes of coming to conclusions about what has been perceived. An individual's type can be measured along four bipolar dimensions: extroversion/introversion; sensing/intuition; thinking/feeling and judgement/perception.

Instrument: A forced-choice, self-report personality inventory which consists of 126 items yielding four scale scores. It is essentially for use with adults and can be administered individually and in groups. Approximate administration time, 50 minutes.

Applications/Implications: Adults may find the type concepts useful for helping to understand basic preferences for learning which can assist in determining compatibility between learning type, method of instruction and other personal or environmental influences on learning.

Canfield Learning Styles Inventory

Definition of Learning Style: Individual learning style is derived from: (a) academic conditions (relations with instructor and peers); (b) structural conditions (organization and detail); (c) achievement conditions (goal setting, competition); (d) content (numbers, words, etc); mode of preferred learning (listening, reading, iconic and direct experience); and (f) expectation of performance level (superior through satisfactory).

Instrument: A self-report instrument based on rank ordering of choices for each of 30 questions. For use with junior high through adult levels. Approximate administration time, 15 minutes.

Applications/Implications: Its major use is to develop instructional materials for whole classes or individual students. The LSI is considered a tool to aid in understanding students' difficulties in completing academic units and for counseling. Emphasis is placed on attitudinal and affective dimensions and the Inventory focuses on such applications.

Gregorc Style Delineator

Definition of Learning Style: Learning style consists of distinctive, observable behaviors that provide clues to the functioning of people's minds and how they relate to the world. These "mind" qualities suggest that people learn in combinations of dualities: (a) concrete-sequential; (b) concrete-random; (c) abstract-sequential; and/or (d) abstract-random. Preferences for a particular set constitute a learning style.

Figure 1. continued,

Instrument: A self-report instrument based on a rank ordering of four words in each of 10 sets. Observation and interviews suggested that these words can be used to aid in categorizing learning preference patterns or modes. For use with upper junior high students through adults. Approximate administration time, 5 minutes.

Applications/Implications: Strong emphasis is placed on the matching of instructional materials and methods to meet the range of individual preferences. Gregorc also recommends that selected nonpreferences be utilized at time to encourage students to strengthen those areas.

Kolb Learning Style Inventory

Definition of Learning Style: Learning style is a result of hereditary equipment, past experience, and the demands of the present environment combining to produce individual orientations that give differential emphasis to the four basic learning modes postulated in experiential learning theory: Concrete Experience (CE); Reflective Observation (RO); Abstract Conceptualization (AC); and Active Experimentation (AE).

Instrument: A self-report instrument based on a rank ordering of four possible words in each of nine different sets. Each word represents one of four learning modes: feeling (CE); watching (RO); thinking (AC); doing (AE). For use with adults. Approximate administration time, 5-10 minutes.

Applications/Implications: Emphasis is placed on individual awareness of personal learning style and available alternative modes. Knowledge of learning style differences should encourage the design of instructional experiences to enhance individual strengths and develop non-dominant orientation.

---

[1] The information in this figure is adapted in part from Dunn, DeBello, Brennan and Murrain, 1981.

THE MYERS-BRIGGS TYPE INDICATOR
Isabel Briggs Myers

The Myers-Briggs Type Indicator (MBTI) is a forced-choice, self-report personality inventory which was developed to measure variables in Carl Jung's theory of psychological type. The MBTI consists of four scales: Extraversion-Introversion (E-I), Sensation-Intuition (S-I), Thinking-Feeling (T-F), and Judgement-Perception (J-P). The most recent version of the Indicator (Form G) was introduced in 1977. It consists of 126 items and is essentially a shortened version of Form F which is also still in use. Of the 40 items eliminated from Form F, 38 were considered experimental and had not been scored on any of the standard scales. Most of the research cited in this review is based on results from Form F.

Practical Features of the Test

Administration

The Myers-Briggs is essentially self-administering. A complete set of easy-to-follow instructions are given on the front page of the test booklet. The directions include instructions on guessing and procedures for marking answer sheets. The answer sheet contains additional instructions for completing the Indicator and includes a sample question to help clarify how the form should be completed.

The Indicator is easily adapted for group administration. The examiner is encouraged to read the directions aloud while the testees read them silently. The instructions given in the test manual, on the test booklet and on the answer sheet are stated clearly enough to insure standardized testing procedures.

Scoring

The MBTI can be hand scored or processed by computer on a dual-purpose answer sheet. Answer keys for hand-scoring are easy to use and contain

---

[1]An updated and revised manual for the MBTI was published in 1985. Consequently, many of the critisims in this critique may not ne applicable to the new manual.

explicit instructions. Since a complete set of instructions is printed on each key, scoring can begin with any key and proceed in any order. Tables for converting raw scores to linearly derived preference scores are also printed on each answer key for ease of scoring.

Other Considerations

Form G takes approximately 50 minutes to complete which is a reasonable amount of time to spend on this type of test. The fact that the Indicator is untimed and non-threatening (testees are told that there are no right or wrong answers) also makes this an attractive test to use with individuals in high school through adulthood. The vocabulary level of the items should not present any difficulty for persons who can read at the high school level and microcomputer software is also available to aid in the interpretation of test results.

Characteristics of the Manual

Reporting of Information

Publishers of the MBTI have provided a comprehensive manual (Myers, 1962) which includes most of the essential information needed for the proper administration and interpretation of the Indicator. Chapters include a detailed description of the purpose of the test, administration and scoring procedures and some suggestions for its potential use. The manual is well written and easy to understand. Individuals with only a basic background in test and measurement techniques and psychology should have little difficulty properly administering and scoring the test.

Parts two and three of the manual provide an overview of the theoretical foundation of the Indicator, a description of how the test items were developed and a well organized presentation on how to interpret the results. While Jung's theory of psychological type includes several abstract concepts,

Myers has done an excellent job of describing the essential characteristics of his theory.

Test Interpretation

At first glance, procedures for interpreting the test results are deceptively simple. The author provides encapsulated descriptions of each of the 16 personality types identified by the four scales. Reviewing these thumbnail sketches can quickly provide a general description of each personality type. However, the test user who wishes to go beyond a superficial interpretation of the results can get quickly lost in the terminology used, the numerous examples cited and alternative ways of interpreting the scores.

Technical Information Provided

Perhaps the most bothersome characteristic of the test manual is the fact that it has undergone very little revision since 1962. A supplementary manual published in 1977 was intended to provide some current normative data but provides a very limited amount of information concerning the reliability and validity of the most recent version of the test (Form G). Only a limited amount of technical data is presented and the the authors seem to be content with stating that "the validity of items does not appear to have diminished" (Myers, 1977, p. 1). This forces the reader either to except this conclusion on faith or search the literature for corroborating information. The paucity of information in this eight page supplement is particularly disappointing after viewing the original manual. Certainly, the extensive use of the Indicator over two decades warrants a complete revision of the manual. At the very least the publishers could include a listing of other sources of information about the technical aspects of the instrument or how it can be most properly used.

Characteristics of the Test

Norms. Normative data for Form F is based on a very substantial number of Massachusetts high school students (N=1872) and an even larger number of liberal arts and engineering students (N=4806) attending several colleges and universities in the Eastern part of the United States. Percentile norms are provided separately for males and females for each of the eight preference scores. Percentile distributions are also provided separately for two groups of high school students (vocational and college bound) and two groups of college students (liberal arts and engineering). The manual also provides frequency distributions for the 16 personality types among students in ten selected fields of study.

The standardization samples used to establish the norms are very loosely defined. The high school students comprising the normative sample were simply described as twelfth grade students from Massachusetts high schools. While the group was separated into academic and vocational groups, there is no indication how the differentiation was made. In addition, there is no description of the methods of sampling used and no identifying information (i.e. locale, socioeconomic status, ethnic background) beyond sex.

The college standardization sample was also poorly defined. All the students were freshmen and all but 240 were males. Although no additional demographic characteristics are provided, the institutions from which the sample originated were prestigious institutions which in all probability biased the sample in favor of students who were above average in intellectual ability and socioeconomic status.

The manual's supplement (Myers, 1977) indicates that, in 1975, Form G was administered to 2,225 children in grades four through twelve "to ensure that cultural changes had not eroded the validity of the Type Indicator" (p. 1).

Again, there is no indication of how the sample was selected. It does appear, however, that this group was biased toward bright (mean I.Q. 117) students who were above average in socioeconomic status. No attempt was made to provide a revised set of percentile norms for this standardization sample and there was no statistical information to determine whether the norms continue to be valid. There was also no indication of differences or similarities for subgroups across ages or grades.

Reliability

The Indicator yields two kinds of scores, dichotomous personality type categories and continuous "preference" scores. Reliability information from the manual along with other research is summarized below and organized according to the kind of score and the aspect of reliability being examined.

Test-Retest Reliability of Type Categories. Test-retest data have been reported using intervals of up to six years. The proportion of individuals who retested into the same type classification ranged from 62% to 90% on each of the four scales (Webb, 1964). Carlyn (1977) summarizes four studies involving college students and a group of elementary school teachers. She reports that in each case, the proportion of agreement between the original and the retest type classifications "was significantly higher than would be expected by chance" (p. 465).

Split-Half Reliability of Type Categories. Essentially three procedures have been used to measure the internal consistency of the four MBTI type categories. Myers (1962) and Webb (1964) report phi coefficients ranging from the low .50's to the high .70's. The samples consisted of both high school and college students and there were no significant differences between the two groups. Lower-bound reliability estimates calculated with Guttman's procedures (Stricker and Ross, 1964) generally yielded lower scores. The

largest coefficient was .73, but most were in the .40's and .50's. Tetrachoric coefficients (Myers, 1962) were generally in the .70's and .80's.

Because the Guttman lower-bound estimates are difficult to interpret without the upper-bound reliabilities (Carlyn, 1977) and tetrachoric coefficients are calculated on the assumption that that the scores are distributed normally (Nunnally, 1978, p. 136) it appears tnat the phi coefficients provide the best estimate of internal consistency reliability for the four type scales. The correlations obtained are somewhat lower than is desirable (Anastasi, 1968, p. 78) for reliability coefficients particularly the T-F scale which is the least consistent.

Test-Retest Reliability of Continuous Scores. There have been surprisingly few test-retest reliability studies for continuous MBTI scores. Stricker and Ross (1964) tested 41 male college students using a fourteen month test interval. Pearson correlations ranged from .48 to .73 across the four indices with Thinking-Feeling yielding the lowest coefficient. Levy, Murphy and Carlson (1972) tested a large group (287 females and 146 males) of Black college students using a two month test-retest interval. Estimates of reliability, also based on Pearson correlations, ranged from .69 to .83. A more recent study (Carskadon, 1977) examined 134 college students with an interval of eight weeks between testing sessions. The coefficients ranged from .56 to .87 and tended to be higher for females than males. The Thinking-Feeling index was the least stable, particularly for males.

In general, the test-retest reliabilities for the MBTI continuous scores are satisfactory although less than optimal for a test of personality traits (Anastasi, 1968). There is a need for additional long range studies with larger populations. In addition, future research should pay particular attention to the Thinking-Feeling scale to determine if it should be revised.

Split-Half Reliability of Continuous Scores. The reliability of continuous scores are somewhat higher than estimates based on the dichotomous type categories. Myers (1962) and Webb (1964) computed product-moment correlation coefficients which produced estimates in the .70's and .80's with a low estimate of .44 for the T-F scale. Stricker and Ross (1964) report similar findings using Coefficient Alpha. Reliability coefficients were generally in the .70's and low .80's. and the T-F scale had lower reliability than the other scales.

Considering these findings, the internal consistency reliabilities appear to be like those of similar self-report inventories (Mendelsohn, 1965) with the exception of the T-F scale which appears the least stable.

Validity

A brief overview of what researchers have to say about three types of validity are described below including content validity, construct validity and predictive validity.

Content Validity. Several researchers have found support for the content validity of the MBTI (Myers, 1962; Carlson and Levy, 1973). In particular, Myers (1962) offers a great deal of evidence for its content validity by citing the methods and criteria used to develop the MBTI's items. Stricker and Ross (1964) examined the content of each item of the four MBTI indices and concluded that the S-N scale and the T-F scale seemed to be consistent with their conceptual definitions. However, the J-P scales and, to a lesser extent, E-I scale seem to be measuring something different than what was intended based on the conceptual definition. Other authors (Coan, 1978; Mendelsohn, 1965; Ross, 1966) support this contention and feel that the E-I and J-P scales measure only limited aspects of the underlying constructs.

Bradway (1964) took a direct approach to content validation by having

Jungian analysts type themselves and then compare their results with the classification produced by three of the four scales on the MBTI. There was fairly high and significant agreement between the two forms of classification demonstrating that the Indicator was valid for the sample of Jungian analysts. Additional evidence for content validity was obtained by Cohen, Cohen and Cross (1981) who used spouses' judgement in predicting the four type scores. Three of these scales, E-I, S-N and T-F received support. The J-P scale failed to show significant agreement between spouses ratings and classification arising from subjective responses on the MBTI.

Although there does not appear to be conclusive support, it appears that the E-I, S-N and T-F scales are generally consistent with Jung's typological theory and the conceptual definitions presented in the MBTI manual (Myers, 1962). If users of the MBTI interpret the J-P scale with caution, the evidence suggests that the test does tap the characteristics the test purports to measure.

Construct Validity. The construct validity of the MBTI has been investigated in numerous correlational studies comparing the Indicator's scores with scores on other instruments. In a series of studies, Stricker and Ross (1964) and Myers (1962) investigated the four scales' correlations with several ability and personality tests. Correlations between MBTI scores and scores from conceptually comparable scales on other instruments were typically in the .60's and .70's providing strong support for the construct validity of the scales. Scores on three of the four ability measures also correlated significantly with the MBTI scales in the predicted direction, however coefficients generally fell in the .10's and .20's.

Additional studies have focused on substantiating the construct validity of a particular scale. Several researchers (Steel and Kelly, 1976; Wakefield,

Sasek, Brubaker & Friedman, 1976) found that the E-I scale of the MBTI correlated positively with the extraversion scale of the Eysenck Personality Questionnaire. The construct validity of the sensing-intuition scale was also supported by a study (Carskadon & Knudson, 1978) which found that as individuals decreased in their preference for concreteness, they were more likely to be classified as an intuitive type on the MBTI.

Taken as a whole, the evidence gathered from a variety of sources presents a strong argument that the scales are measuring the attitudes formulated by Jung and conceptualized by Myers.

Predictive Validity. With regard to predictive validity, research has focused on career and achievement related variables. The MBTI has been shown to be moderately predictive of success in a physician-extender training program (Buhmeyer & Johnson, 1978) and job satisfaction among pediatric nurse practitioners (Bruhn, Bunce and Floyd, 1980). Other reports indicate that there is evidence to suggest that the MBTI can contribute to the prediction of retention of college students and that they "relate meaningly (sic) to a large number of variables including personality, ability, interest, value, aptitude and performance measures, academic choice, and behavior ratings" (Mendelsohn, 1965, p. 322).

Overall Quality of the MBTI

The test author and publisher have made a concerted effort to develop an evaluation tool which (1) approaches personality assessment from a nonpathological point of view, (2) produces results which are easy to apply and (3) provides information which describes the way people view and interpret the world around them. I feel that the Indicator has succeeded in producing a mechanically well developed instrument which most individuals would find interesting and non-threatening.

The manual published in 1962 is quite extensive and contains a very thorough description of the theoretical foundations and methods for interpreting the test results. However, the manual presents a limited amount of information to support the reliability and validity of the test. Examiners interested in using the Indicator to assess the personality types of secondary and postsecondary students would be hard pressed to find a sufficient amount of information to support its use.

The samples used to establish normative data for Forms F and G are very restricted and poorly defined. As a result it is difficult to determine which individuals can be appropriately compared with the data reported. Both split-half and test-retest reliability coefficients tend to be somewhat low for this type of instrument. The Thinking-Feeling scale appears to be particularly unstable, suggesting that much more research is needed to determine which extraneous factors are influencing this score. While there is a considerable amount of information to support the content and construct validity of the Indicator, the question of whether it is effectively tapping the Jungian constructs underlying the test has not been conclusively established.

Personal Decision Regarding Use of the Myers-Briggs

Despite its shortcomings, I consider the Myers-Briggs to be one of the better instruments currently available to assess learning style type. However, while the Indicator appears to be a good instrument in terms of its theoretical and empirical bases I would be reluctant to use it in lieu of other instruments which provide more direct measures of aptitude, career interests, satisfaction, etc. At the present time too little is known about how Myers-Briggs constructs can be applied to assist an individual with

educational and career decisions. Until the test can be validated using a more representative sample of adolescents and adults I feel the test should be used for facilitating discussions of learning style type and research purposes only. More information is needed before the Indicator's results can be used reliably and validly with individuals to make predictions about career choice, interests or preferred learning style.

LEARNING STYLE INVENTORY
David Kolb

On the basis of his model of experiential learning, David Kolb developed the Learning Style Inventory (LSI). This self-administered questionnaire consists of nine word sets. Each set has four words for a total of 36 word choices. Examinees are asked to rank the four words according to how well each word characterizes his or her individual learning style. Twenty-four of the 36 words are related to one of the four learning style dimensions: abstract conceptualization (AC), concrete experience (CE), active experimentation (AE), and reflective observation (RO). Twelve additional words are included as distractors.

Two additional composite scores are computed from the learning style dimensions: (a) the relative amount of abstractness or concreteness in learning style (AC-CE) and (b) the relative degree of activeness or reflectiveness (AE-RO). These two difference scores place an individual in one of the four quadrants formed by the intersections of the AC-CE and AE-RO axes. A dominant learning type is identified according to the learning style preferred: accommodator, diverger, converger and assimilator.

Practical Features of the Test

Administration

The Learning Style Inventory is designed to be self-administering. Individuals interested in taking the test are given a self-scoring test and interpretation booklet which includes instructions on how to complete, score and interpret the test results. The LSI is completed by ranking nine sets of four words that are the best and least "characteristic of you as a learner". The LSI can be administered individually or in groups. The format is attractive and easy to follow but can be easily modified to include only the instructions and the test protocol. Tests can then be scored later by the examiner.

Scoring

The LSI is usually scored by hand in a section of the test booklet directly below the nine word sets. The word sets are arranged in four columns of nine words each. Each column represents one of the four learning style dimensions. Each of the LSI scales is based on the sum of the ranking of six

words in each column. Three words in each column serve as distractors.

A complete set of instructions is printed in each test booklet. Boxes are provided to record individual and total scores for each preference. Computation of the two combination scores is also simple and straightforward using the format provided.

Other Considerations

The LSI is untimed, but generally takes about 10 minutes to complete. This makes it an attractive test to use for both guidance and research purposes.

The format and approach of the LSI provides a very non-threatening "environment" for the evaluation of learning style. Examinees are reminded that there are no right or wrong answers and that the purpose of the inventory is to describe the individual's learning style, not to evaluate learning ability. The vocabulary level is designed for individuals in their late teens and should present little difficulty for the average adult. However, there is some indication that individuals with low levels of academic achievement may have difficulty understanding the meaning of some of the words (Posey, 1984).

A final consideration is that the measurement format of the LSI requires that the instrument be classified as an ipsative measure (Anastasi, 1968). Ipsative scores are designed to assess the relative strength of each learning style in relation to the individual's other learning style preferences. As a result, the scores of one individual can not be compared with those received by someone else. Consequently, individuals with the same learning style type (i. e. accommodator, converger, diverger, assimilator) may differ markedly in the absolute strength of their learning styles. The use of ipsative scales in the Kolb LSI also raises some questions regarding the appropriateness of statistical analyses which are typically performed on normative data

(Anastasi, 1968; Merritt and Marshall, 1984b).

Characteristics of the Manual

Reporting of Information

Kolb (1976) has prepared a technical manual which attempts to cover most of the basic requirements established by the American Psychological Association (1974) for appropriate test development. The first two sections of the manual provide a very thorough description of the purpose of the LSI and the experiential learning theory upon which it is based. The reader interested in experiencial learning theory will find the manual's treatment concise and understandable.

Chapter III includes a description of the internal properties of the test including item analysis, intercorrelations of the LSI scales, reliability and descriptive statistics. Validity information is reviewed in the fourth chapter which focuses primarily on predictive and construct validity information. The final sections include a bibliography of references using the LSI and an appendix containing information on a normative sample.

Test Interpretation

Although a description of the each of the four learning style types can be found in the test manual (Kolb, 1976, p. 5-6), most of the interpretive information can also be found in the test booklet. The information consists of a brief description of each of the learning modes - concrete experience (CE), reflective observation (RO), abstract conceptualization (AC), and active experimentation (AE) and each learning style type - assimilator, accommodator, converger, diverger. Two methods of interpreting an individual's scores are provided. The first approach is to plot the raw scores from the four different learning modes on a graph resembling a target. The concentric

circles which comprise the "target" represent the approximate percentile scores of the normative group described in the manual. Either method can be used to determine how an individual's scores compare with the percentile scores of the normative group. However, as indicated earlier, the ipsative nature of the scores makes this a highly suspect procedure. The strength of a particular score is influenced by the strength of the other three. Therefore, the only appropriate norm would be between the individual scores. A more appropriate interpretation could consist of having the individual who completes the LSI list the four learning modes in order of strength (i. e. highest raw score) to determine which learning style they prefer the most, then second, etc.

The second interpretive approach seems to make the most sense both from a practical standpoint and from a psychometric point-of-view. An explanation is provided for calculating the two comparison scores (AC-CE and AE-RO). These scores are plotted on a grid with a single horizontal and vertical axis. By marking their raw scores for these two scales on the grid at their point of intersection all individuals can determine their dominant learning style as either an accommodator, diverger, converger or assimilator. A summary of the four basic learning style types is contained on the final page of the booklet. According to Kolb, these descriptions are based on both research and clinical observation of these patterns of LSI scores.

Technical Information Provided

The technical manual for the LSI was originally published in 1976 and revised in 1978. The LSI was created by a panel of "behavioral scientists" who were familiar with Kolb's experiential learning theory. An explanation of how the instrument was developed is included along with a description of the of the intercorrelations between the LSI scales.

Information concerning the reliability of the test includes the reporting of test-retest and split-half reliability coefficients. The samples used consisted exclusively of full time graduate students or students returning to school to complete their graduate work. Descriptions of the background characteristics, testing conditions, etc. under which the tests were administered are limited.

The normative samples for the LSI include a group of management students and adult norms derived from a diverse group of individuals but consisting primarily of college students. The management group includes five groups of management students from Harvard and M.I.T. The adult norms are based on a combination of 13 groups of adults and the management groups described above.

The validly section of the manual is undoubtedly the weakest from a psychometric perspective. The information provided consists almost exclusively of construct validity information and much of the conclusions drawn are speculative. The studies examined the relationship between the LSI and performance tests, personality tests, teacher preferences for learning situations and academic specialization. The methodology used and descriptions of the samples are very limited. In particular, the studies which focused on preferences for learning situations and academic specialization provided very little information to judge the validity of the results. Consequently, it is difficult to draw any firm conclusions regarding the usefulness of the test, based on the information provided in the manual.

Characteristics of the Test

Normative Information

Normative data for the LSI consists of essentially two groups of individuals. The first group is comprised of five different sample groups

of men "who are involved in managerial careers" including graduate students from Harvard and M.I.T., Sloan Fellows, who come to M.I.T. for one year to complete a master's degree in management and two groups of active managers. The total sample consists of 741 people. Generalized adult norms are also reported. This normative group consists of 18 group samples including the management groups described above, college undergraduates, graduate students and several professional occupation samples.

The normative information provided is disappointing for three reasons. First, both normative groups are definitely biased toward the upper ranges of general intellectual ability, socioeconomic status and levels of education when compared to the general population. This would make comparison of scores questionable when considering high school students, adults with average or below average levels of intellectual ability and individuals with limited formal education.

Secondly, although the manual indicates that there are sex and age differences on the LSI, no separate norms are provided across these characteristics. The normative information provided is limited to the means and standard deviations of each group for each of the six scores (the four scales plus the two composite scores). As a result, the two norms tables represent a composite of scores from the groups described.

Finally, the norms tables themselves provide only approximations of the corresponding percentile score for a particular raw score. The tables are divided into deciles with the raw scores for each scale located between the lines representing the decile points. Consequently, it is difficult to determine accurately what percentile rank corresponds to a particular raw score unless it happens to fall directly on a line.

Reliability

There is a paucity of information in the professional literature concerning the stability and consistency of LSI scores. In addition to the information reported in the technical manual, only four articles (Freedman & Stumpf, 1978; Geller, 1979; Merritt, 1983; Merritt and Marshall, 1984b) could be found in professional journals. One of these articles (Merritt, 1983) simply stated that the coefficients ranged from .52 to .89 but did not provide a breakdown of reliability estimates for each of the scales. Essentially three different types of reliability estimates are discussed including two measures of internal consistency and test-retest reliability. The estimates across studies are fairly consistent and are summarized below.

Internal Consistency Reliability. Estimates of reliability based on coefficients of internal consistency were calculated using a variety of

Table 1. Internal Consistency Reliability Coefficients for the Learning Style Inventory

Reference	Sample	n	CE	RO	AC	AE	AC-CE	AE-RO
Kolb, 1976	MIT Sloan Fellows	47	.69	.37	.65	.64	.78	.78
Kolb, 1976	MIT Sloan Fellows	50	.43	.59	.81	.61	.80	.81
Kolb, 1976	Active Managers	90	.61	.58	.71	.62	.78	.85
Kolb, 1976	Harvard MBA's	442	.50	.63	.74	.67	.75	.84
Kolb, 1976	Lesley Undergrads	58	.48	.63	.74	.65	.82	.86
Freedman & Stumpf, 1978	Business Grad Stud	412	.33	.61	.69	.51	.71	.72
Freedman & Stumpf, 1978	Business Grad Stud	1179	.40	.57	.70	.47	.71	.66
Merritt & Marshall, 1984	Nursing Students	187	.29	.59	.52	.40	-	-

student populations. Table 1 shows split-half reliabilities obtained from studies reported by Kolb (1976) and two additional studies which have appeared

in the professional literature. The coefficients reported by Kolb (1976) are Spearman-Brown split-half reliability coefficients. Freedman and Stumpf (1978, 1980) and Merritt and Marshall (1984b) used Coefficient Alpha. The split-half method provides a measure of consistency with regard to content sampling. Coefficient alpha has the advantage of taking into account not only the content sampling, but also the heterogeneity of the behavior domain sampled.

The internal consistency of the instrument as a whole is relatively low. The split-half coefficients are comparable for all five samples, where the concrete experience scale (CE) is the least reliable ($\bar{x}$ = .54) and the abstract conceptualization scale (AC) is the most reliable ($\bar{x}$ = .73). The difference scales (AC-CE, AE-RO) have moderate reliability for the five samples with average correlation coefficients of .79 and .83 respectively.

The alpha coefficients of .29 to .71 reported were consistently lower than the Spearman-Brown reliabilities. This could be due in part to the fact that Kolb (1976) made a conscious effort to divide the test so that the items which correlated most highly were placed in alternate halves. As a result, some of the heterogeneity of the test was artificially controlled using this particular split-half method. Despite the generally lower Alpha coefficients the overall pattern of results remained the same. The concrete experience scale (CE) had the lowest average reliability ($\bar{x}$ = .34) and abstract conceptualization (AC) had the highest ($\bar{x}$ = .70). The difference scales were also estimated to be more reliable than the individual scales but demonstrated only moderate reliability.

Test-Retest Reliability. Test-retest reliabilities range from .34 to .73 with intervals ranging from 31 days to seven months (See table 2.). These reliability estimates are fairly low ($\bar{x}$ = .53) suggesting that an individual's

ranking of the words is not particularly stable over time. Of the thirty-six correlations listed, only four are in the .70's; 17 are in the .50's and .60's and 15 are in the .30's and .40's. The average reliability estimates among the four scales and the difference scores were fairly consistent, ranging from .48 to .61. The abstract conceptualization score (AC) demonstrated the highest reliability while the concrete experience scale (CE) appeared to be the least stable.

It should also be noted that the group samples used to estimate the test-retest reliability consisted exclusively of students in business management and medicine. Thus the question of comparability to other groups of individuals becomes an important interpretation issue.

Validity

Evidence Presented in the Manual. The first reference to the validity of the LSI is found in the item analysis section of the technical manual. Intercorrelations between the words that comprise the four scales are described and generally correlate in the expected directions. Kolb (1976) concludes: "This data shows that the words comprising the four primary LSI scales have both high convergent and discriminant validity." (p. 10) However, the Standards for Educational & Psychological Tests (American Psychological Association, 1974) states: "Correlations of item scores with total scores on the test in which the item is included (or a parallel form of that test) may be presented as item-discrimination coefficients, but they should not be presented or used as item-validity coefficients." (p. 32) The Standards booklet further points out that these data are useful for thinking about construct validity but that they are indicators of internal consistency, not validity.

Section IV of the manual, which discusses the validity of the LSI, reviews

Table 2. Test-Retest Reliability Coefficients for the Learning Style Inventory

Reference	Sample	Interval	n	CE	RO	AC	AE	AC-CE	AE-RO
Kolb, 1976	Medical Students	3 mos.	27	.48	.73	.64	.64	.61	.71
Kolb, 1976	MIT Grad Students	3 mos.	23	.48	.51	.73	.43	.51	.48
Kolb, 1976	MIT Grad Students	6 mos.	18	.46	.34	.64	.50	.53	.51
Kolb, 1976	MIT Sloan Fellows	7 mos.	42	.49	.40	.40	.33	.30	.43
Geller, 1979	Medical Students	31 days	50	.56	.52	.59	.61	.70	.55
Freedman & Stumpf, 1978	Business Grad Stds	5 weeks	101	.39	.49	.63	.47	.58	.51

several correlational studies relating the LSI scores to performance tests, personality tests, academic specialization, and preference for learning situations and teachers. The evidence presented is equivocal and in some instances actually yields inconsistent results.

Correlations between LSI scores and five performance tests (the ATGSB, which is not described, the Law School Admissions Test, the Wunderlic Aptitude Test, the Remote Associates Test and the Uses of Objects Test) were seldom significant. Only three of the 48 correlation coefficients listed were above .30 indicating very little shared variance between measures. In addition, correlations between the performance test scores and LSI scale scores were not consistent across different types of students.

An examination of correlations between the LSI and scores on the Myers-Briggs, the Thematic Apperception Test and the Firo-B also provided little support for the construct validity of the Inventory. A comparison of LSI scores with the Myers-Briggs seemed to support the LSI constructs but not consistently in all groups. No relationships between the LSI and the Thematic Apperception Test or the Firo-B were hypothesized. In both instances, however, only a few of the correlations were statistically significant and none of them exceeded the .40 level.

Additional studies, all of which were completed by Kolb, examined the relationship between the LSI scales and the preferences of 144 Harvard MBA's for teachers and learning situations and learning styles and academic specialization which was based on a sample of 800 practicing managers and graduate students in management. By-and-large, most of the conclusions and interpretations drawn from these data are based on "appearances" and conjecture and are not substantiated by the statistical analysis of the data.

Prediction of Career Choice. Most of the research which has used the LSI

as a predictor of career choice has involved individuals in medical professions. Plovnick (1975) administered the LSI to 47 medical students in an attempt to determine whether students with different learning styles were attracted to specific career choices within the medical field. He concluded that there was an association between type of medical career chosen and specific learning styles.

Wunderlick and Gjerde (1978) conducted a replication study involving 172 practicing physicians and resident physicians and 44 medical students. They criticized Plovnick's original investigation because of the small sample size (n=47), the lack of statistical analyses and a failure to classify individuals correctly into the four learning style types. Statistical analysis of their data did not support a relationship between learning style and medical career choice. They concluded: "for the purpose of discriminating learning style differences among career groups it appears necessary to construct a new instrument" (p. 54) and recommended that the LSI not be used to provide career guidance to medical students.

Four additional studies in the medical literature support the use of the LSI but they are largely anecdotal. Sadler, Plovnick and Snope (1978) surveyed family practice physicians and medical faculty and report a percentage distribution of the four learning style categories within the two groups. Approximately 50 nurse practitioners were asked by Christensen, Lee and Bigg (1979) to complete the LSI near the end of their professional training. While 70% of the group fell in either the accommodator or diverger category, no differences in performance was observed between any combination of the four learning style types.

The other two anecdotal studies involving individuals in the medical profession include a study by Leonard and Harris (1979) who used the LSI with

a small group of residents and staff in an internal medicine residency program and Baker and Marks (1981) who conducted a learning style analysis of 21 anesthesiologists. Both articles give qualified support to the LSI but no statistical evidence is presented.

More elaborate efforts to study the relationship between the LSI and career choice are reported by West (1982) and Merritt (1983). West (1982) concluded on the basis of his findings that there was no consistent relationship between the personality traits described in the LSI manual and the traits measured by the Myers-Briggs Type Indicator and the Omnibus Personality Inventory. He further concludes that the LSI may not be effective in explaining individual learning styles within the medical professions and that more validity studies of the LSI are needed.

Likewise, when Merritt (1983) studied the LSI scores of nearly 500 RN students, she found no relationship between age, work experiences and learning preferences and the learning style categories.

Prediction of Performance in Educational Settings. In general, studies which have examined the relationship between the LSI and specific instructional methods have also been anecdotal. Whitney and Caplan (1978) compared the LSI results of a group of family practice physicians who completed a refresher course and a group who had not attended the course. No predominant learning style type emerged and there were no significant differences between the two groups. However, the authors did give qualified support to the idea that individuals prefer a specific type of instruction which is compatible with their preferred learning style .

A large sample (n=503) of college juniors and seniors enrolled in a principles of management course were randomly placed in laboratory sections which emphasized either discussion, an experiential mode of instruction or

simulation. The results do provide some support for matching learning style type with method of instruction. However, some of the results were not statistically significant and there were some inconsistencies in the data which the authors admitted were difficult to explain. "The results seem to indicate that learning style is a useful tool in curriculum development at the university level. It appears that students might reach higher levels of academic performance if learning style is used as an aid in individualizing learning environments" (Brenenstuhl and Catalanello, 1979, p. 29).

Fox (1984) studied the relationship between different learning styles as measured by the LSI and participants' evaluations of a specific program. No relationship was found, leading Fox to seriously question the construct validity of the LSI. He also found no association between learning styles and reactions to different methods of instruction. He concludes that "without further validation of the relationship between the LSI and either learner preferences or learner performance, one must question the usefulness of the LSI as a guide to educational design decisions" (p. 84).

Two studies employed the LSI in an attempt to predict levels of performance in courses with computer based instruction. Reifle and Edwards (1975) found no significant differences in students' preferences for learning and various computer based instructional techniques. Descriptive statistics and correlations were used by Kevin and Liberty (1975) to compare computer based and traditional instruction in a chemistry course. The findings conflicted with the hypothesized correlation between major and the LSI. However, as predicted, the concrete-experience scale of the LSI did correlate positively with grade.

Pigg, Busch and Lacy (1980) investigated the relationship between the LSI and implications for designing education programs using a group of county

extension agents from Kentucky. The results failed to support the idea that there is a specific relationship between learning styles abilities and preferences toward specific instructional techniques.

Studies Using Factor Analysis. A limited number of studies have examined the structure of the LSI using factor analysis techniques. Ferrell (1983) found that when items comprising the scale loaded on four primary factors which generally matched the four learning styles described by Kolb. The factor loadings accounted for about one third of the variance in scores. She concluded that the results tended to support two bipolar learning style dimensions but that further work was necessary to improve the psychometric properties of the instruments.

Lamb and Certo (1978) compared LSI results using both the original inventory and a seven point Likert scale. They found the LSI provided results equivalent to previous research. The modified instrument produced different results. They concluded that the support for learning style theory may be due to instrument bias.

In a follow-up study Certo and Lamb (1979) randomly generated responses to the LSI using a Monte Carlo technique. After the statistical analysis of the data provided some support for the learning style theory, they concluded that the design of the LSI spuriously supports its theoretical base. They further concluded that "the use of the theory to make normative judgements about educational practices should be suspended until the above problem is rectified" (p. 447).

Freedman and Stumpf (1980) examined the average LSI scores for different undergraduate majors and found that less than five percent of the between group variance could be accounted for by learning style. A factor analysis of the data also provided weak support for the two bipolar dimensions theorized

by Kolb. Like Certo and Lamb (1978, 1979), the authors concluded that "much of the accounted for variance may be a function of the ipsative scoring system used with the LSI. Because the four scales are interdependent, high scores on one dimension force lower scores on the other dimensions" (p. 446).

Overall Quality of the LSI

The Learning Style Inventory is an attempt by Kolb to operationalize his experiencial learning theory and provide a normative assessment of preferred learning style. As in all measures of hypothetical constructs, the reliability and validity of the instrument is critical.

Kolb argues that traditional forms of assessing reliability may not apply to the LSI due to the "interdependent (i.e., any action, including responding to the test, is determined in varying degrees by all four learning modes) and variable (i.e., the person's interpretation of the situation should to some degree influence which mode he uses)" (Kolb, 1976, p. 12) nature of the characteristics measured by the test. Nevertheless, he does provide test-retest and split-half reliability coefficients for several groups of students. Additional reliability information can be found in the professional literature on other groups of students.

From the above analysis of the available reliability data it appears that the LSI yields rather unstable scores. With the exception of the combination scores (AC-CE and AE-RO) which are higher, the remaining correlation coefficients are only moderately reliable and fall in a range which are generally not exceptable for measures which are assessing hypothesized constructs (Anastasi, 1968, p. 78; Nunnally, 1978, p. 245). These low reliabilities limit the ability of the inventory to explicate learning styles (Freedman and Stumpf, 1978, p.280). Finally, it appears that the Inventory "will be of limited use for assessment and selection of individuals" (Kolb,

1976, p. 13) and is probably unsatisfactory for differentiating among individuals or between large disparate groups (Geller, 1979).

In addition to the low reliability estimates the evidence reviewed suggests that both the construct and predictive validity of the LSI has not been confirmed. Studies have attempted to verify the validly of the LSI by identifying factors e.g., career choice, preferred instructional method, college major, personality characteristics, and so on, which should theoretically correlate with specific learning styles. In nearly every instance, where statistical analyses were performed, the results were equivocal and inconsistent.

Factor analytic studies have also provided questionable support for the construct validity of the LSI. Perhaps the most revealing studies were those conducted by Certo and Lamb (1978, 1979) and Freedman and Stumpf (1980) who presented fairly convincing evidence that the construction of the instrument may be confounding the results, since the four scores are derived by ranking only two independent dimensions.

Personal Decision Regarding Use of the LSI

The LSI has been used extensively in management education, medical education and most recently has been applied to numerous adult and continuing education situations. In many educational applications, I suspect that the ability of the LSI to accurately identify preferred learning style or basic personality characteristics is never called into question. However, the information reviewed here does seem to raise some serious doubts about the appropriate use of the Inventory. While both the reliability and validity of the LSI is in question, several authors have suggested that the evidence does provide support for the learning model itself (Fox, 1985; Merritt & Marshall,

1984; Pigg, Busch and Lacy, 1980). Pigg, et al (1980) go so far as to say:

> "Despite these cautions against utilizing inventories such as Kolb's for developing educational programs, the Learning Style Inventory does appear to be a useful instrument. A number of individuals, including these researchers, have reported that the inventory really captured the tendencies in their personal behavior. Being able to recognize these tendencies, and relate them to behavior patterns is important. Thus, it is concluded that the LSI may be effectively employed as a useful device in the actual conduct of educational programs or in a participatory approach to the development of adult education programs due to its high degree of face validity." (pp. 242-243)

However, this appeal, which seems to be based on the premise that if the instrument seems to work well we don't need to worry about its psychometric quality, places the LSI on very shaky grounds. In the opinion of this reviewer, the unreliability and lack of evidence for either construct and predictive validity suggests that the LSI could produce very misleading results and needs to be studied much more carefully before it should be used in any setting.

LEARNING STYLES INVENTORY
Alfred A. Canfield

The Learning Styles Inventory is a self-report measure of learning style that is concerned with determining selected attitudinal values people have toward the teaching-learning situation. The inventory consists of 30 items each followed by four possible responses. The respondent is asked to rank (on a scale of one to four with one being the most descriptive) the responses according to how well they describe their personal reaction or feelings. Twenty scale scores are derived from the responses to the items which fall into basically four areas: conditions, content, mode, and expectancy.

Practical Features of the Test

Administration

The Learning Styles Inventory is self-paced and designed primarily for use with adults, however, the manual does include norms for junior and senior high school students. The inventory can be administered individually or in small or large groups. There is no specific time limit, but the manual indicates that completion time generally ranges from 20 to 45 minutes. It is recommended that at least 50 minutes be set aside for administering the inventory to groups of 30 or more.

The test can be easily administered and scored by individuals with only a limited amount of training and experience in test administration procedures. Test score interpretation can be completed successfully by individuals who have taken the time to read the descriptive information contained in the test manual. Individuals who have had graduate level training in a professional area (e.g. psychology, counseling, psychiatry, tests and measurements) would be better prepared to incorporate the results of the inventory with other relevant information. However, this level of expertise does not appear necessary to make adequate use of the inventory's results. Instructions provided in the manual are not standardized but they appear simple and clear

enough to ensure consistent test administrations.

Scoring

At present the Inventory can only be hand-scored. Under the appropriate circumstances, respondents can score their own protocols. The hand-scoring procedure is fairly efficient and is completed directly on the test protocol. The calculation of the "Overall Expectancy Score" is a bit complicated but can be mastered after a couple of dry runs. Twenty-one scale scores including the overall expectancy score are calculated for each individual. Profile forms for graphing the the scale results are also available. In addition to raw scores only percentiles scores are provided using the tables included in the manual. Instructions for plotting percentile scores from the norms tables are also provided in the manual.

Other Considerations

The reading level of the inventory appears to be low enough for students in high school, however, a cursory examination of the items suggests that a junior high school student may have a difficult time understanding the meaning of a number of the items. I also suspect that high school students would need very good reading skills in order to fully comprehend the intent of many of the questions. The content of many of the items (some refer to final exams, turning in a paper to an instructor and teacher training) strongly suggests that the Inventory is geared for adults.

Another interesting aspect of the test is that many of the items require the respondent to imagine a hypothetical situation. If the individual has never experienced the situation described, it may be difficult for some individuals to develop the appropriate mind set to respond to the item the way the author of the Inventory intended (e.g. Question 13 asks the respondent to imagine that they are required to visit a home for the elderly).

Characteristics of the Manual

Reporting of Information

The manual (Canfield, 1980) includes an adequate description of the what the Inventory is intended to measure. However, there are some glaring omissions of a number of essential test manual elements established by the American Psychological Association (1974). For example, the manual does not include a description of how the inventory was developed and standardized, there is no description of the normative groups and very limited information about the reliability and validity of the scale. These omissions leave the potential user with little evidence to judge the strengths and weaknesses of the Inventory.

Test Interpretation

Interpretation of the Inventory's results is based on percentile scores derived from the norms presented in the manual. The manual recommends using a system in which "key" scores are derived based on preset percentile score levels. Percentile scores are then classified as "very strong", "strong", "middle", "low" or "very low". Individual scores or group summaries can then be interpreted by focusing on those scales which fall in the "strong" or "very strong" categories. A set of directions is provided outlining interpretation procedures for group data. The implicit assumption seems to be that individual score profiles can be interpreted in the same way.

The back side of the profile sheet consists of a brief description of the scales. The manual includes 21 pages of text describing the learning preferences of individuals who score high on a particular scale. As a result, test administrators who wish to go beyond the brief summary description must wade through a large amount of information to make any sense out of the scores. This section concludes with several listings of instructional

techniques related to the four modes measured on the Learning Styles Inventory.

My general reaction to this entire section of the manual is that it is poorly organized, extremely difficult to understand and seems to encourage the user to restrict interpretation of results to the brief summary on the back of the profile form. A great deal of effort on the part of the evaluator would be required to use the interpretative information provided.

Characteristics of the Test

Norms

Descriptive information pertaining to the normative samples used to calculate percentile scores is virtually nonexistant. The manual includes six separate norms tables in the back of the manual: male norms, female norms, high school male and female norms and junior high male and female norms. The only other information provided is the number of individuals included in each sample. The general male and female norms are based on a sample of 1,364 and 1,180 individuals respectively. The high school and junior high school male and female norms are based on samples of approximately 100 students in each group.

This paucity of information regarding the standardization sample is particularly bothersome because the test user has no way of determining whether the scores being interpreted can be appropriately compared with the normative population. Although few test manuals contain all of the information deemed "essential" by the American Psychological Association (1974) Canfield's manual includes basically none of the elements. For example, there is no indication of when the normative data was gathered, the population is not defined and the method of sampling is not discussed. There is no description of such relevant variables as ethnic status, socioeconomic

level, age, sex, locale and educational attainment. Finally, the manual fails to provide basic descriptive statistical information in regard to the normative group including measures of central tendency and variability.

Reliability

The reliability information provided in the test manual is also extremely limited. Canfield reports that a set of scale reliabilities, based on a sample of 369 community college students, were calculated utilizing the "Froelich method".[1] He reports that the reliabilities ranged from .59 to .92 but provides no additional information. This suggests that some scales have high reliability while others have fairly low reliability. Unfortunately there is no way of determining which scales fall into which categories. No test-retest reliability is reported.

The only additional reliability information reported in the manual is a set of split-half reliability coefficients "supplied by Dr. Steve Brainard and Dr. Jerry Omen of Longview Community College, Lee's Summit, Missouri" (Canfield, 1980, p. 51). The coefficients listed all range in the very high .90s which normally would be outstanding! However, the fact that these results are not supported by other researchers suggests that more research is needed before any firm conclusions can be drawn.

A study completed by Merritt and Marshall (1984a) contains the only other reliability information I was able to locate in the professional literature. They report estimates of the internal consistency reliabilities using Coefficient Alpha. The reliabilities ranged from .54 to .82 based on a sample of 187 nursing students. Of the sixteen coefficients reported, three

---

[1] I could not find this method described in Anastasi, 1968 or in Nunnally, 1978.

fell in the low .80s, seven were in the .70s, three were in the .60s and three were in the fifties. Generally speaking, Alpha coefficients which fall below .80 suggest a moderate to low amount of internal consistency.

In short, reliability data for the Learning Styles Inventory is sorely lacking. The information which is available suggests that many of the scale scores are highly volatile and could change dramatically from administration to administration.

Validity

Support for the validity of Canfield's Inventory, as reported in the test manual, is limited to a description of differences among program majors at a community college in Missouri. The studies were conducted in 1976 by Brainard and Osmen (as reported by Canfield, 1980). Eight groups of students were compared: secretarial students, data processing students, females enrolled in a special development program, enlisted men in the military, students in an Art History course, "educationally disadvantaged" veterans, and community college teachers. In most instances, the narrative suggest that the scores received by these groups were in the predicted direction. However, the manual provides absolutely no statistical data to support the results of the study.

The validity section of the manual concludes with a description of "all studies known to have been completed by January 1, 1980" (Canfield, 1980, p. 65). It describes, in some detail, a study which establishes cutoff scores to differentiate between achieving and non-achieving students. The study is based on slightly more than one hundred students. There is no reference provided to obtain additional identifying information. The statistical data presented is limited to means and t-values.

The remaining references include two articles published in refereed

journals, four doctoral dissertations and two unpublished papers. Limited information is provided about the results of these studies and no statistical information is provided.

Factorial Validity. Merritt and Marshall (1984a) completed a study of 187 nursing students who were administered the Learning Styles Inventory as part of larger study. A factor analysis was used to identify factors measured by the test. The factor structure yielded eight identifiable factors rather than the twenty identified by Canfield. The authors conclude that the subscales defined by Canfield within each major section of the instrument do not form independent useable factors. They suggest that the model should be collapsed to reflect the factors identified in the study.

Overall Quality of the Learning Styles Inventory

The Learning Styles Inventory provides a self-reported measure of how individuals feel about various aspects of a learning environment. The revised 1980 manual has a very limited amount of technical information making it difficult to adequately judge the quality of this test. The inventory is easy to administer individually or in groups and, at least on a superficial level, provides information which counselors, teachers and school administrators can readily understand.

Realibility information reported in the test manual and the professional literature is wholly inadequate to make any judgements about whether the inventory provides a stable measure of learning style. Correspondingly, the lack of available evidence regarding the scale's validity makes any interpretation of the Inventory scores highly suspect.

Another glaring weakness of the scale is the normative percentile scores provided in the manual. There is absolutely no description of the composition

of the normative groups or when the data was collected. As a result the user of the Inventory has no way of knowing whether the standardization sample is appropriate for their intended use. All things considered the manual is very inadequate in terms of the standards for educational and psychological tests published by the American Psychological Association (1974).

Personal Decision Regarding the Use of the Canfield LSI

In my opinion, the only redeeming aspect of the Learning Styles Inventory is its face validity. The description of the scales developed by Canfield appear to be potentially useful to educators and administrators in adult education who are seeking ways to better match a learner's preferences for a particular learning environment with an instructional method. The single published study which reports reliability coefficients suggests that some of the scales may be reliable. The split-half reliabilities reported in the manual are spuriously high and suspect do the limited sample and small number of items comprising each scale. The only information concerning the test's validity suggests that there may be some relationship between the Inventory subtest scores and a students choice of major. Much more research needs to be conducted, however, before I would feel comfortable using the subscale scores for this or any other purpose.

In summary, because there is so little information available regarding the psychometric characteristics of the test, including information pertaining to reliability and validity. Consequently, if used it at all, the LSI should be for research purposes only. The manual is so inadequate and there is so little psychometric information available in the professional literature, the inventory should probably be described as only an experimental assessment tool.

GREGORC STYLE DELINEATOR
Anthony Gregorc

The Gregorc Style Delineator was designed to be a self-administered, self-analysis tool. The scale consists of ten sets of four words. Individuals are asked to rank the words that are the most and least descriptive of themselves (a four would indicate the most powerful descriptor, while a one would indicate the least powerful descriptor). The Delineator yields scores in four categorical areas: Concrete random, concrete sequential, abstract sequential, abstract random. Each category score has a possible range of 10 to 40 and is based on the sum of the rankings of 10 words. The categories examined by the Delineator are intended to aid the individual in recognizing and identifying the "channels through which he/she receives and expresses information".

Practical Features of the Test

Administration

The Gregorc test protocol is designed for self-administration and self-scoring. Person's interested in completing the Delineator can complete it individually or in groups. As with most other learning styles instruments the Delineator is not timed but the directions recommend about four minutes to complete the ranking of the 40 stimulus words.

The protocol and directions for scoring and graphing the results are reproduced on a single 8 1/2 x 17 sheet of heavy paper stock. The directions for ranking the words and completing the the scoring procedures are straightforward, clearly laid out and easy to follow. High school students and adults should have no difficulty completing and scoring this learning styles instrument.

Scoring

The Gregorc is designed to be scored by hand. The ten word sets are arranged to facilitate the scoring and calculation of the four channel scores. Each score is based on the rankings of ten words. Raw scores for each scale can range from 10 to 40. A style profile is included on the answer

sheet which allows the person completing the test to graphically locate their scores in one of the four quadrants formed by the intersection of the two bipolar dimensions. The author has developed a scoring continuum based on a series of interviews and identified three score ranges: 27-40 - high "pointy-head"; intermediate "moderate" and low "stubby point". Brief synopses of the dominant style characteristics of the four channels are also included on the back of the answer sheet.

Other Considerations

The Delineator is very attractively packaged and is obviously designed for quick administration and scoring. Although there is no indication that the words were formally evaluated in terms of their reading difficulty it appears that they are basic enough to be understood by adults and adolescents who are reading at the high school level. Both nouns and verbs are used.

Another important consideration for the potential user of this learning style tool, is that the words have been arranged so that the words which comprise a scale are all in the same row. This makes it very easy for the individual who is taking the test to determine which words go together. It's possible that after ranking one or two sets of words an individual could consciously or unconsciously bias the results by consistently ranking the words in a particular row either high or low. However, this possibility is not discussed in the Delineator's administration manual.

Characteristics of the Manual

Reporting of Information

The technical and administration manual was published in 1982. The five sections of the booklet contain information about the development of the test and its theoretical base (Section 1), the validity of the delineator (Section

2 & 4), reliability (Section 3), concluding remarks (Section 5) and administration guidelines (Section 6).

A first glance, the manual appears to have all the essential elements established by the American Psychological Association (1974). A closer examination, however, reveals that very little empirical evidence is provided to support the claims of the author. Most of the information provided is based solely on the author's experience and appears to be based on a limited number of studies with small sample sizes.

Test Interpretation

Interpretation of the Delineator's results is based on the total score an individual receives for each of the four mediation channels, concrete sequential, abstract sequential, abstract random and concrete random. After graphing the results on the backside of the Style Delineator, an individual is able to identify their dominant learning style. A synopsis of the characteristics of each type is printed on the form. The administrator is also encouraged to use the publication, An Adult's Guide to Style (Gregorc, 1982) "and appropriate personal experiences to 'flesh-out' the interpretations" (Gregorc, 1984, p. 29).

The introduction section of the manual (Gregorc, 1984) indicates that the graphing of matrix scores was designed to illustrate the bipolar oppositions of the four styles identified by the Delineator. It also states that the graphing "provides the potential for using the Gregorc Style Delineator as an educational psychotherapeutic tool for counselors and advisors" (p.6). However, following a brief statement regarding the interpretation of the results, the manual includes the following disclaimer:

> "The Gregorc Style Delineator is not for diagnosis or prescription; it is designed for self-analysis, for self-observation, and for prompting

understanding of self, others, and environments. An individual must be given the right to self-validate and accept, suspend judgement on, or deny his scores. The instrument 'works' for the vast majority. It does not appear to 'work' for everyone. This fact must be acknowledged in order that results are not used to conveniently or devastatingly label or pigeonhole one's self or another human being." (p. 29-30)

Because of these two somewhat conflicting statements and the limited amount of information in the manual, it appears that individuals should interpret the results with extreme caution. The descriptions of the four primary types were derived from interviews with more than 400 individuals. However, the selection criterion used was simply an individual's willingness to share perceptions. There is virtually no information provided to describe the characteristics of these individuals, leaving the question of how adequately this sample represents a particular individual or group completely unanswered.

Finally, the lack of a psychological or empirical basis for the Delineator makes prior experience with adults and knowledge of adult development totally ineffective in interpretation of the test data.

Characteristics of the Delineator

Norms

One of the most glaring weaknesses of the test manual is that no normative information is provided. The only clue the test user has about the scoring criteria is enmeshed in the description of how the scale was developed.

The stimulus words for the Delineator were borrowed in large part from an instrument, the Transaction Ability Inventory, developed in the 1970's by the author (Gregorc, 1978). Interviews with 40 individuals (no identifying characteristics are provided) and the judgements of 22 graduate students (who

were simply described as being knowledgeable of the theory of mediation ability) determined which words were assigned to each scale. The list of words was then reduced by removing words "which could be considered jargon associated with the educational field" (Gregorc, 1984, p. 7). This was accomplished by polling 60 adults "who were from private industry".

The scoring criteria was then arbitrarily established by dividing the range of scores for each style into three groups. The manual states that the upper group represented scores which fell at or above the 74th percentile. The lower group included individuals with scores below the 27th percentile. These score ranges were then adjusted through a series of personal interviews before final score ranges were determined. Scores in the high group (27 to 40 points) are described as "pointy-head", scores in the middle range (16 to 26 points) as "moderate" and scores in the lower third (10 to 15 points) as "stubby-point".

This complete absence of descriptive and statistical information regarding a normative sample leaves the interpretor of the Delineator's results with virtually no basis for making any interpretations of the raw scores. It appears that Gregorc expects the user simply to accept on faith that the scores and the accompanying descriptions of the four basic learning style types identified by the instrument are valid.

Reliability

The manual includes the results of only one study to support the reliability of the Delineator. No information could be found in the professional literature. The study cited in the manual was conducted by Gregorc (1984) and is based on 110 adults who took the Gregorc Style Delineator on two occasions ranging from six hours to eight weeks.

Measures of internal consistency reliability are provided in the form of

standardized Alpha coefficients which ranged from .89 to .91. Test-retest reliability coefficients ranged from .85 to .88. All of these coefficients suggest that the Delineator is a highly reliabile instrument. However, there are several factors which may have unduly influenced these results, creating spuriously high correlations.

First, Gregorc did not control for differences in test-retest intervals. He simply pooled the data and reported a single reliability coefficient for each scale. There is no indication of how many individuals fell into the six hour category or how many completed the Delineator a second time after an eight week interval. Obviously, one would expect greater stability of scores over shorter time intervals.

Second, the structure of the Delineator's protocol makes it extremely easy for the individual completing the test to "decipher" how the test works. This greatly enhances the probability that, after an individual rates one or two sets of words, a conscious or unconscious effort will be made to rate the rest of the word sets consistently, creating spuriously high Alpha coefficients. In addition, the test-retest reliability coefficients may also be influenced by this factor. It is relatively easy for someone to remember their high and low scores over a six hour to eight week period. This in turn, makes it relatively easy to reproduce practically the same ratings for each of the ten sets of words the second time the Delineator is completed.

In short, the methodological weakness of the study reported in the manual in conjunction with the format of the Delineator's answer sheet suggests that the coefficients may not accurately reflect the internal consistency of the scores or their stability over time. Much more research needs to be completed before any judgements can be made.

Validity

Section 2 of the manual purports to discuss the information available regarding the validity of the Delineator. Two aspects of validity are discussed, construct validity and predictive validity. The evidence presented to support the construct validity of the instrument consists mainly of what Gregorc calls a "definitional" approach to construct validation. In practice, this approach consisted of defining the four constructs two different ways across six pages in the manual. There is no statistical support for the definitions, no "expert" testimony provided and no attempt to relate the definitions to any theory of personality or psychological development. The only statistical data provided to support the construct validity of the instrument are the Alpha coefficients discussed earlier which were based on an undefined sample of 110 adults.

The fourth section of the manual provides a description of studies which purport to measure the predictive validity of the Delineator. However, the author's description of the studies indicates that this research more appropriately falls into the construct validation category. The methodology used in both studies is poorly described but enough information is provided to suggest that the results may be seriously flawed.

In the first study subjects were asked to complete the Delineator and rate themselves on a list of 40 items which were described as representing the four domains measured by the test. Validity coefficients ranging from .55 to .76 were interpreted by the author as providing "moderately strong" support for predictive validity. However, because there is no evidence of the extent to which the criterion itself (i.e. the 40 item test produced for purposes of the study) actually measures the constructs being measured, no firm conclusions can be drawn from the study. In addition, there is not enough information

provided to rule out such contaminating factors as having similar or identical words in both instruments and the order in which the tests were taken.

The second study is even more weak methodologically. After administering the Gregorc Style Delineator to the 475 subjects, each was given a list of characteristics attributed to their classification as yielded by the instrument. Each subject was then asked to indicate to what extent those attributes described him or her on a five point scale. The author reports that 89% of the 475 subjects agreed or strongly agreed that the attributes described them. No additional statistical analyses are reported.

This procedure for validating an assessment tool is much like reading your horoscope in the newspaper at the end of the day. In most instances you can recall at least one situation or event which occurred during the day which corresponds to the prediction made. In fact one could probably mix-up the predictions assigned to the various astrological signs and still get a high level of agreement.

Overall Quality of the Gregorc Style Delineator

The Delineator is described by the author as a self-analysis tool, "specifically designed to aid an individual to recognize and identify the channels through which he/she receives and expresses information efficiently, economically, and effectively" (p. 1). The most attractive features of the Delineator is the "packaging" of the test, the quick administration time, ease of scoring and interpretation of results. However, the quality of the instrument ends there.

A review of the psychometric information provided in the manual provides little information to support the reliability and validity of the instrument. Normative data is nonexistant. The validity and reliability information provided is so limited and methodologically flawed that no firm conclusions

can be drawn from any of the information provided. A review of the professional literature yielded no empirical studies which used the Delineator. Without additional information the only conclusion which can be drawn is that the Delineator, at least from a psychometric point of view is of very poor quality.

Personal Decision Regarding the Use of the Delineator

Because of all the shortcomings described above, the Gregorc Style Delineator appears to have little practical value to the individual seeking a better understanding of their personal learning style. I believe that the most appropriate use of this instrument would be to provide an example of how not to construct a assessment tool. The almost total lack of a theoretical basis for the scale coupled with its questionable reliability and validity eliminates all practical purposes for its use. Until considerably more statistical support for the scale becomes available the instrument should probably be used strictly for research purposes.

SUMMARY, CONCLUSIONS AND SUGGESTIONS FOR FURTHER RESEARCH

Summary

Content, Format and Scoring

A wide range of research studies pertaining to four learning style instruments (i. e. the Myers-Briggs Type Indicator, Kolb Learning Style Inventory, Canfield Learning Styles Inventory and Gregorc Style Delineator) were reviewed to address the issue of whether they are of sufficient psychometric quality to warrant their continued use either for research or educational purposes. The instruments characterize several different theoretical orientations and measure a variety of dimensions typically associated with the learning style concept. The content of each instrument is different and there appears to be very little overlapping of the dimensions measured. However, they do have several features in common.

First, all the instruments are designed to be a self-report measure of learning style. Respondents are asked to indicate or rank their choices by indicating what appeals to them the most. The Myers-Briggs requires respondents to choose between 126 pairs of statements (actually seven items have three options and one has four). The Canfield LSI has 30 items, each with four options, which individuals are asked to rank. The Kolb LSI and Gregorc instruments require respondents to rank sets of four words including nouns and verbs.

Second, the scoring of these measures mainly consists of summing the rankings obtained for each item which comprises a particular scale. The Kolb, Canfield and Gregorc instruments are designed to be self-scoring while in most instances it is more efficient to have someone other than the respondent score the Myers-Briggs. In addition to raw scores, only percentile scores are

generated from an individual's responses. None of the instruments provide standardized scores.

It is also important to keep in mind that all of the instruments reviewed employ ipsative scores, that is, the strength of each learning style category is expressed, not in absolute terms, but in relation to the strength of the respondent's other learning style preferences. Therefore, the proper frame of reference is the individual rather than a normative sample (Anastasi, 1968).

Third, the instruments are also similar in the techniques used for identifying the respondent's learning style profile. Each instrument allows the respondent to plot their results on a chart which will identify predominant learning style preferences. Three of the scales use bipolar dimensions which allow an individual to be placed in a specific type category. The Myers-Briggs yields 16 possible learning style types. The Kolb LSI and Gregorc's Style Delineator identifies four possible categorical areas. The Canfield yields 20 scale scores within four general categories.

Reliability

Studies which have investigated the reliability of the instruments usually report either test-retest or internal consistency coefficients. In general, the test-retest reliabilities for the MBTI are satisfactory although less than optimal for a test of personality traits (Anastasi, 1968). No information reporting test-retest reliability coefficients for the Kolb LSI could be found in the professional literature. Coefficients reported by Kolb (1976) generally range from the low 40's to the high 70's suggesting that the scores are less stable than the Myers-Briggs.

No test-retest information could be found for the Canfield LSI and only one study, conducted by Gregorc (1984), was found in the Delineator's technical manual. Considering the very limited amount of information it

appears that no firm conclusions could be drawn regarding the stability of the scores produced by these instruments.

Internal consistency reliability coefficients for all four instruments were also very limited. In most instances the split-half and Alpha coefficients which were reported fell in the 60's and 70's. These estimates, which are lower than desirable (Anastasi, 1968), suggest a moderate to low amount of internal consistency.

Validity

Construct validity was the most frequently discussed type of validity. Studies of one type or another were reported for all of the instruments and several different methods were used. For example, studies using the MBTI and Kolb LSI correlated scores with individuals' educational specialization, career choice and current job. In addition, the Kolb LSI has been used a great deal in predicting career choice in medical and business settings. Scores from the Kolb LSI and MBTI have also been used in several factor analytic studies and, in some instances, have been compared with scores from instruments that measure similar dimensions or constructs.

In general, the data provide equivocal support for the validity of these instruments. However, in some instances, studies which used the Myers-Briggs did result in a relatively acceptable degree of construct validity. Most of the studies cited by Kolb (1976) had weak methodology and poorly defined research samples. The limited amount of validity information provided for the Canfield LSI and Gregorc's Delineator was so limited and methodological flawed that no firm conclusions could be drawn from the information provided.

Conclusions

To anyone familiar with the field of adult education it is obvious that

there is a growing number of instruments available for assessing learning style. I believe that much of this growing interest can be traced to a strong desire, on the part of adult educators, to meet the needs of a very diverse group of learners. Proponents of the learning style concept (e. g. Cross, 1976, Keefe, 1979 and Smith, 1982) feel that learning style is a viable concept with important implications for both adult educators and learners. These implications include the possibility of achieving a better understanding of oneself as a learner and help with facilitating the learning of others.

At the present time, learning style instruments are being used to facilitate career planning (Kolb, 1984; Torbit, 1981), diagnose learning difficulties (LeFlar, 1982) and make decisions about teaching and helping people learn (Chiarelott & Davidman, 1983; Dunn, 1984; Dunn, Dunn & Price, 1981; Gregorc, 1979). Since results may influence students' career plans or attitudes toward learning, it seems particularly important to pay more serious attention to the psychometric quality of the instruments being used. Poor quality learning style instruments could be generating data that are weak or misleading. As Freedman and Stumpf (1978) aptly point out, "Measurement error remains measurement error no matter how effectively an exercise or instrument is applied within a class" (p. 281).

From the preceding review of literature it seems apparent that there are significant measurement and related technical problems present in all of the instruments reviewed. First, none of the instruments have established an appropriate normative base for the valid interpretation of scores. At the very least, each of these measures should have a well defined sample of adult continuing education students including percentile distributions by sex and age. Without these reference points any interpretation of scores becomes

highly suspect.

Secondly, there appears to be an incomplete development of many of the theoretical constructs underlying the instruments reviewed. Evidence supporting the construct validity of the Myers-Briggs is minimal and practically nonexistent for the Kolb LSI, the Canfield LSI and the Gregorc Style Delineator. The few factor analytic studies which were completed with the MBTI and Kolb LSI vary in the degree of their support for the constructs which are supposed to be measured. This suggests that either there is a problem with the construction of the instruments or the learning style paradigm is lacking. I suspect that there are problems with both. Therefore, studies which contribute to the construct or predictive validation of these instruments are sorely needed.

Third, estimates of reliability provided by the research reviewed for this paper suggest that learning style preferences are somewhat unstable, even for relatively short periods of time. Nevertheless, it could be argued that the dynamic nature of learning style makes high test-retest reliability coefficients unnecessary and that a greater emphasis should be placed on the homogeneity of the instrument from a single administration. However, in general, the reported Alpha coefficients, which measure this characteristic, also suggest that the scores produced may not be reliable indicators of learning style preference.

Finally, the ipsative scores produced by these instruments appear to be influencing the results of many of the validity studies which appear in the literature. Factor analytic studies, in particular, strongly suggest that the construct validity of the bipolar conceptualizations of learning style are artificially supported by this type of ranking procedure. As a result, studies which produce normative scores from the same items or newly developed

items may provide stronger evidence for the validity of a scale. In addition, providing a normative basis for the scores may also make interpretation of the results easier. The combination of normative and ipsative frames of reference currently provided in the test manuals makes the interpretation of scores very difficult and less meaningful than would be the case with a consistently ipsative or consistently normative approach.

Suggestions for Future Research

Despite the criticisms and difficulties described above, serious consideration should continue to be given to the measurement of the learning style construct. A large number of questions concerning the assessment of learning styles require further attention. Among these it is suggested that future research look into the following areas of inquiry.

1. Little research, using these four instruments has assessed the relationship between actual learning and either learning styles or preferences for particular instructional techniques. Research related to these issues would not only contribute to the existing body of knowledge pertaining to construct validity but would also add to our knowledge of learning style as an instructional tool.
2. Research results to date suggest that no learning style measure by itself provides an adequate basis for an individual to select a career or be counseled to do so. Consequently, future research could focus on such questions as: Are different career categories characterized by specific learning style types? Are people whose career and learning style match more successful or satisfied? How much weight should be given to learning style and career choice?

3. A concerted effort should be made to establish large representative norms for returning adult students. This could include normative samples from groups of credit and noncredit college and university students, distance learners and individuals engaged in informal learning projects.
4. More information should be gathered relative to the match between learning styles and the environment. This could involve comparing scores with objective measures of achievement, reports by learners on their choice of instructional method, observations of behavior, results from projective tests and reported satisfaction with the instructional environment.
5. Future research could also explore the issue of whether an individual's preferred learning style is modified by the educational environment. For example, do adult students learn better when instruction is adapted to their learning style preferences? Can people be trained to adopt a particular learning style? Do learning styles remain stable over time in adult populations? Does a significant change in life situations result in changes in learning style?

It seems apparent that a valid model and measurement device for learning style would be a powerful tool for the facilitation of learning. The specific requirements for the optimal learning style instrument have also been carefully outlined by Grasha (1983) and would:

- "demonstrate internal consistency and test-retest reliability;
- exhibit construct and predictive validity;
- produce data that can be translated into instructional practices;
- produce high degrees of satisfaction among learners placed in environments on the basis of the information it provided;

- help facilitate learners' ability to use content; and
- perform its magic in ways that are clearly superior to those possible without it" (pp. 30-31).

I'm sure that the authors of the learning style instruments reviewed for this paper would accept these criteria. However, while some of the instruments show promise in meeting some of these criteria, none of them can claim to have met them all.

REFERENCES

American Psychological Association. (1974). Standards for educational and psychological tests. New York: American Psychological Association.

Anastasi, A. (1968). Psychological Testing (Third Edition). Toronto: Macmillan Company.

Baker, J. D. & Marks, W. E. (1981). Learning style analysis in anesthesia education. Anesthesiology Review, 8(7), 31-34.

Bradway, K. (1964). Jungs psychological types: Classification by test versus classification by self. Journal of Analytical Psychology, 9, 129-134.

Brenenstuhl, D. C. & Catalenello, R. F. (1979). Using learning styles to improve academic effectiveness. Journal of Experiential Learning and Simulation, 1, 29-37.

Bruhn, J. G., Bunce, H. III, & Floyd, C. S. (1980). Correlates of job satisfaction among pediatric nurse practitioners. Psychological Reports, 46, 807-814.

Buhmeyer, K. J. & Johnson, A. H. (1978). Predicting success in a physician-extender training program. Psychological Reports, 42, 507-513.

Canfield, A. A. (1980). Learning styles inventory: Guide. Birmingham, Michigan: Humanics.

Canfield, A. A. & Lafferty, S. C. (1974). Manual for the learning styles inventory. Plymouth, Michigan: Experiential Learning Methods.

Carlson, R. & Levy, N. (1973). Studies of jungian typology: memory, social perception and social action. Journal of Personality, 41, 559-576.

Carlyn, M. (1977). An assessment of the Myers-Briggs Type Indicator. Journal of Personality Assessment, 40, 655-662.

Carskadon, T. G. (1977). Test-retest reliabilities of continuous scores on the Myers-Briggs Type Indicator. Psychological Reports, 41, 1011-1012.

Carskadon, T. G. & Knudson, M. L. (1978). Relationship between conceptual systems and psychological types. Psychological Reports, 42, 483-486.

Certo, S. C. & Lamb, S. W. (1980). An investigation of bias within the learning styles inventory through factor analysis. Journal of Experiential Learning and Simulation, 2(1), 1-7.

Chiarelott, L. & Davidman, L. (1983). Learning styles inventories: Implications for curriculum and instruction. U. S. Educational Resources Information Center, ERIC Document ED 231 840.

Christensen, M. G., Lee, C. A. & Bugg, P. W. (1979). Professional development of nurse practitioners as a function of motivational need, learning style and locus of control. Nursing Research, 28(1), 1979, 51-56.

Claxton, C. S. & Ralston, Y. (1978). Learning styles: Their impact on teaching and administration. AAHE/ERIC Higher Education Research Report No. 10. Washington, D.C.: American Association for Higher Education.

Coan, R. W. (1978). Review of the Myers-Briggs Type Inventory. In O. K. Buros (Ed.) Eighth Mental Measurements Yearbook. Highland Park, N.J.: Gryphon Press.

Cohen, D., Cohen, M. & Cross, H. A. (1981). A construct validity study of the Myers-Briggs Type Indicator. Educational and Psychological Measurement, 41, 883-891.

Cross, K. P. (1976). Accent on learning. San Francisco: Jossey-Bass.

Dunn, R. (1981). Madison prep: Alternative to teenage disaster. Educational Leadership, 38(5), 286-387.

Dunn, R. (1984). Learning style: State of the science. Theory into Practice, 23(1), 10-19.

Dunn, R., Debello, T., Brennan, P., & Murrain, P. (1981). Learning style researchers define differences differently. Educational Leadership, 38(5), 372-375.

Dunn R., Dunn, K. & Price, G. E. (1981). Learning styles: Research and opinion. Phi Delta Kappan, 62(9), 645-646.

Ferrell, B. G. (1983). A factor analytic comparison of four learning styles instruments. Journal of Educational Psychology, 75(1), 33-39.

Fox, R. D. (1984). Learning styles and institutional preferences in continuing education for health professionals: A validity study of the LSI. Adult Education Quarterly, 35(2), 72-85.

Freedman, R. D. and Stumpf, S. A. (1978). What can one learn from the Learning Style Inventory? Academy of Management Journal, 21(2), 275-282.

Freedman, R. D. and Stumpf, S. A. (1980). Learning style theory: Less than meets the eye. Academy of Management Review, 5, 445-447.

Geller, L. M. (1979). Reliability of the Learning Style Inventory. Psychological Reports, 44, 555-561.

Grasha, A. (1983). Learning style: The journey from Greenwich Observatory (1796) to Dalhousie University. In L. Curry (Ed.). Learning Style in Continuing Medical Education. Ottawa, Ontario: Council on Medical Education, Canadian Medical Association.

Gregorc, A. F. (1978). Transaction ability inventory. Columbia, CT.

Gregorc, A. F. (1979). Learning-teaching styles: Potent focus behind them. Educational Leadership, 36, 234-236.

Gregorc, A. F. (1982). An adults guide to style. Maynard: Mass.: Gabriel Systems Inc.

Gregorc, A. F. (1984). Gregorc style delineator: Development, Technical and administrative manual. Maynard, Mass.: Gabrial Systems, Inc.

Hunsaker, J. S. (1980). The experiential learning model and learning style inventory: An assessment of current findings. Journal of Experiential Learning and Simulation, 2(3), 145-152.

Jenkins, J.M. (1982). Teaching to individual student learning styles. The Administrator, 6(1), 10-12.

Kagan, J. & Kogan, N. (1970). Individual variation in cognitive processes. In P. H. Mussen (Ed.). Carmichael's manual of child psychology, Volume 1. New York: Wiley.

Keefe, J. W. (1979). School applications of the learning style concept: Student learning styles. Reston, Va.: National Association of Secondary School Principals.

Kevin, R. C. & Liberty, P. G. (1975) Student's personality, attitude and learning styles as predictors of performance in undergraduate organic chemistry course using computer based education. U. S. Educational Resources Information Center, ERIC Document ED 115 209.

Kolb, D. A. (1976). Learning Style Inventory: Technical Manual. Boston: McBer.

Kolb, D. A. (1984). Experiential Learning. Englewood Cliffs, N. J.: Prentice-Hall.

Kolb, D. A. & Fry, R. (1975). Toward an applied theory of experiential learning. In C. L. Cooper (Ed). Theories of Group Processes. London: John Wiley.

Lamb, S. & Certo, S. (1978). The learning style inventory (LSI) and instrument bias. Academy of Management Proceedings, 28-33.

Leflar, S. M. (1982). Implementing a learning style inventory. Journal of Development and Remedial Education, 5(Special Issue), 4-5.

Leonard, A. & Harris, I. (1979). Learning in a primary care internal medicine residency program. Archives of Internal Medicine, 139, 872-875.

Levy, N., Murphy, C. & Carlson, R. (1972). Personality types among Negro college students. Educational and Psychological Measurements, 32, 641-653.

Mendelsohn, G. A. (1965). Review of the Myers-Briggs Type Indicator. In O. K. Buros (Ed.) Sixth Mental Measurements Yearbook. Highland Park, N. J.: Gryphon Press.

Merritt, S. L. (1983). Learning style preferences of baccalaureate nursing students. Nursing Research, 32(6), 367-372. (Photocopy)

Merritt, S. L. & Marshall, J. C. (1984). Reliability and construct validity of alternate forms of the CLS inventory. Advances in Nursing Science, 7(1), 78-85.

Merritt, S. L. & Marshall, J. C. (1984). Reliability and construct validity of ipsative and normative forms of the Learning Style Inventory. Educational and Psychological Measurement, 44(2), 463-472.

Messick, S. (1984). The nature of cognitive styles: Problems and promise in educational practice. Educational Psychologist, 19(2), 59-74.

Myers, I. B. (1962). The Myers-Briggs Type Indicator. Palo Alto: Consulting Psychologists Press.

Myers, I. B. (1977). Supplementary manual: The Myers-Briggs Type Indicator. Palo Alto, California: Consulting Psychologists Press.

Nunnally, J. C. (1978). Psychometric Theory. New York: McGraw Hill.

O'Donnel, J. P., Paulson, K. A., & McGann, J. D. (1978). Matching Familiar Figures Test: A unidimensional measure of reflection-impulsivity? Perceptual and Motor Skills, 47, 1247-1253.

Pigg, K. E., Busch, L. & Lacey, W. B. (1980). Learning styles in adult education: A study of county extension agents. Adult Education, 30(4), 233-244.

Pizzo, J. (1982). Learning styles: Teaching through small group techniques. Early Years, 12(9), 32-35.

Plovnick, M. S. (1975). Primary career choices and medical student learning styles. Journal of Medical Education, 50, 849-855.

Posey, E. J. (1984). Learning style inventory: Implementation research. Journal of Developmental and Remedial Education, 7(3), 16-18.

Reifle, S. & Edwards, A. (1975). A C-BE education instrument: Descriptive report. U. S. Educational Resources Information Center, ERIC Document ED 115 210.

Ross, J. (1966). The relationship between a Jungian personality inventory and tests of ability, personality and interest. Austrialian Journal of Psychology, 18, 1-17.

Sadler, G. R., Plovnich, M. & Snope, F. C. (1978). Learning styles and teaching implications. Journal of Medical Education, 53, 847-849.

Smith, R. M. (1982). Learning how to learn: Applied theory for adults. New York: Cambridge.

Steele, R. S. & Kelly, T. J. (1976). Eysenck personality questionnaire and Jungian Myers-Briggs Type Indicator correlation of extraversion-introversion. Journal of Consulting and Clinical Psychology, 44, 690-691.

Stricker, L. J. & Ross, J. (1964). An assessment of some structural properties of the Jungian personality typology. Journal of Abnormal and Social Psychology, 68, 62-71.

Torbit, G. (1981). Counselor learning style: A variable in career choice. Canadian Counselor, 15(4), 193-197.

Wakefield, J. A., Jr., Sasek, J., Brubaker, M. L. & Friedman, A. F. (1976). Validity study of the Eysenck Personality Questionnaire. Psychological Reports, 39, 115-120.

Webb, S. C. (1964). An analysis of the scoring system of the Myers-Briggs Type Indicator. Educational and Psychological Measurement, 24, 765-781.

West, R. I. (1982). A construct validity study of Kolb's learning style types in medical education. Journal of Medical Education, 57(10), pt. 1, 795-796.

Whitney, M. A. & Caplan, R. M. (1978). Learning style and instructional preferences of family physicians. Journal of Medical Education, 53, 684-686.

Witkin, H. A. & Goodenough, D. R. (1981). Cognitive styles: Essence and origins - field dependence and field independence. New York: International Universities Press.

Wunderlick, R. & Gjierdl, C. L. (1978). Another look at the learning style inventory and medical career choice. Journal of Medical Education, 53, 45-54.